see abc of epidemiology on page 411

Measuring disease occurrence

KJ Jager[1], C Zoccali[2], R Kramar[3] and FW Dekker[1,4]

[1]ERA–EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [2]CNR–IBIM Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Renal and Transplantation Unit Ospedali Riuniti, Reggio Cal., Italy; [3]Austrian Dialysis and Transplant Registry (OEDTR), General Hospital of Wels, Wels, Austria and [4]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands

Different measures may be used to describe how often disease (or another health event) occurs in a population. Incidence expresses the development of new cases and is mostly used against the background of prevention, to assess disease etiology or to determine the risk factors of disease. Depending on the specific study question, incidence may be reported as risk or as incidence rate. This paper discusses that it is preferable to use incidence rate in case of a dynamic population or in cases where the observation period is sufficiently long for competing risks or loss to follow-up to play a significant role. Prevalence is the number of existing cases, which is affected by both the number of incident cases and the length of disease time. It reflects the burden of disease on a population that may, among others, be measured in terms of costs or morbidity. Knowledge about this burden can be used for the planning of health-care facilities. This paper discusses the different measures of disease occurrence using a number of examples taken from the nephrology literature.

Kidney International (2007) **72,** 412–415; doi:10.1038/sj.ki.5002341; published online 20 June 2007

KEYWORDS: disease occurrence; risk; incidence rate; prevalence

Correspondence: *KJ Jager, ERA-EDTA Registry, Academic Medical Center, Department of Medical Informatics, J1b 115.1, PO Box 22700, 1100 DE Amsterdam, The Netherlands. E-mail k.j.jager@amc.uva.nl*

Received 2 March 2007; revised 29 March 2007; accepted 3 April 2007; published online 20 June 2007

Epidemiology is the study of the occurrence of disease. In this case 'disease' should be interpreted quite broadly, as epidemiology studies many types of health outcomes or events. The Centers for Disease Control and Prevention therefore use a wider definition like 'the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems'.[1] The type of measure of disease occurrence to be used for analysis depends on the purpose of a study.

If we are performing a study against the background of prevention and we aim to assess the etiology of a disease or event and determine its risk factors, we are interested in the development of *new cases* of that disease over a period of follow-up, the so-called *incidence*. Should we, on the other hand, wish to know the burden of disease on a population because we need this for the planning of health-care facilities it is much more useful to know the number of *existing cases* that is expressed by the *prevalence* of disease.

INCIDENCE

Two measures of disease occurrence deal with new cases: risk and incidence rate (for a definition of terms see Table 1). Risk is a proportion; it is the ratio of the number of subjects developing disease (or other health outcome) over a specific period to the number of subjects followed:

$$\text{Risk} = \frac{\text{Number of subjects developing disease during a time period}}{\text{Number of subjects followed for the time period}}$$

To quantify *risk* (synonyms: cumulative incidence, incidence proportion), it is always necessary to define a time period to which the risk applies. This can simply be illustrated with the concept of risk of death. We as humans can be fairly certain that the risk of death within 150 years is 100%, whereas the risk of death within 1 day will usually be quite small. Secondly, the concept of risk assumes that subjects are followed for the entire time period. That such may not always be the case is illustrated by example 1 that was taken from the paper of Puliyanda *et al.*[2]

Example 1 – Risk

The paper of Puliyanda et al.[2] *describes a cohort of 3106 children during the first 2 years post-renal transplantation. One*

of the purposes of the study was to determine the risk of hospitalization for bacterial infection in the first 2 years after renal transplantation. One hundred and sixty-four children lost their grafts in the first 6 months after transplantation. Six hundred and eighty-seven patients were hospitalized for bacterial infection.

In this example, what would be the risk of hospitalization for bacterial infection in the first 2 years post-renal transplantation? There were 3106 children 'at risk' at the moment of transplantation. As 687 children developed bacterial infection for which they needed to be admitted to hospital the risk should, according to its definition, be calculated as $687/3106 = 22.1\%$. The problem was, however, that 164 children lost their graft for another reason than bacterial infection and were therefore not able anymore to develop the event of interest. This example shows that there are problems with the concept of risk as a measure of disease occurrence. In general, such problems will occur if the observation period is relatively long and study participants may cease to be at risk for the event of interest, for example, because they die from other causes or get lost to follow-up.[3] One could consider the risk of death from other causes as 'competing' with the risk of the event of interest. Although, intuitively, risk is relatively easy to interpret, in cases where the observation period is sufficiently long for competing risks or loss to follow-up to play a significant role, risk may be less suitable as a measure of incidence.[4] As explained in the next paragraph, other circumstances where risk may not be the preferred measure of disease occurrence include the use of dynamic populations as populations at risk and in studies where events can happen more than once in one individual. Therefore, in many cases it is better to use incidence rate.

Incidence rate is the ratio of the number of subjects developing disease (or other health outcome) to the time at risk for disease:

$$\text{Incidence rate} = \frac{\text{Number of subjects developing disease}}{\text{Total time at risk for the subjects followed}}$$

This formula shows that incidence rate differs from risk in that the denominator includes a measure of time instead of a number of subjects. In this perspective, incidence rate is an instantaneous concept, like speed. A major advantage of using incidence rate (synonyms: incidence density, hazard, force of morbidity/mortality) compared to using risk is that it is not required for every study subject to complete the entire risk period, as only 'time at risk' is taken into account. This property makes the incidence rate very useful in dynamic populations, in cases where subjects may or may not be at risk for the event of interest for particular periods of time. Suppose we would be interested in the incidence rate of peritonitis requiring hospitalization in continuous ambulatory peritoneal dialysis (CAPD) patients in 2004 and we would have diagnosed a number of such peritonitis episodes in 17 CAPD patients. We would then need to calculate the total time at risk, that is, the total time on CAPD. Figure 1 shows that together these 17 patients were 144 patient-months at risk. In this period, there were four of such episodes. The incidence rate of peritonitis episodes requiring hospitalization would therefore be $4/144 = 0.028$ per patient-month or $4/12 = 0.33$ per patient-year.

Example 2 – Incidence rate

For the year 2005, Kramar and Oberbauer[5] reported a number of 374 renal transplants in an Austrian population of 8.1 million inhabitants. The incidence rate of renal transplantation in Austria in that year was therefore 46 transplants per million person-years or, as it is usually stated, per million population.

Another practical application of incidence rate is renal transplant rate as shown in example 2. In order to be able to compare transplant activity between countries, registries divide annual transplant numbers by the number of country inhabitants. For a dynamic population as the general population, it is unfeasible to calculate the different times at risk for different persons and then add them up. However, under steady-state conditions persons dying in this general population are being replaced by newborns and, therefore,

Table 1 | Definitons of terms

Concept		Definition	Formula
Incidence	Risk	Probability of developing disease	No. of subjects developing disease during a time period / No. of subjects followed for the time period
	Incidence rate	Ratio of the number of cases to the time at risk for disease	No. of subjects developing disease / Total time at risk for the subjects followed
Prevalence	Point prevalence	Proportion of people in a population having disease at a particular point in time	No. of subjects having disease at a particular point in time / Total no. of subjects in the population
	Period prevalence	Proportion of people in a population having disease over a period of time	No. of subjects with disease at the start of the period + no. of subjects developing disease over the time period / Total no. of subjects in the population

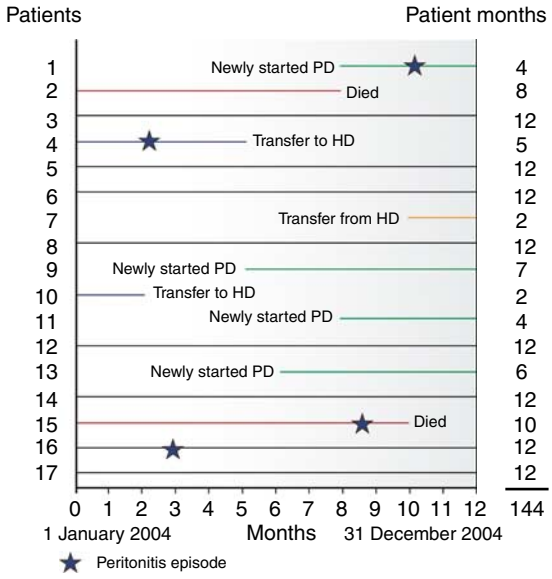

Figure 1 | Time at risk for peritonitis requiring hospitalization in 17 CAPD patients.

the population number can be taken as a proxy for the number of person-years lived in the general population.

Another property of incidence rate is that, under steady-state conditions in which rates do not change with time, it is the reciprocal of the waiting time, that is, the average time until an event occurs. If we go back to the example of the incidence rate of peritonitis requiring hospitalization in CAPD patients, we could easily calculate that the average waiting time for a peritonitis requiring hospitalization to occur in a CAPD patient is $1/0.33 = 3$ years.

One can imagine that when measured over short periods of time, the risk and the incidence rate will be similar, as there will be little loss to follow-up and competing risk will only play a minor role. For this reason, in-hospital mortality for acute renal failure is commonly expressed as a risk, whereas mortality on dialysis for end-stage renal disease (ESRD) is usually expressed as a rate. A further comparison of the properties of risk and incidence rate is shown in Table 2.

Table 2 | A comparison of properties of risk and incidence rate (adapted from Rothman[4])

Property	Risk	Incidence rate
Range	0–1 (0–100%)	0– infinity
Units	None	1/time
Interpretation	Probability	Reciprocal of waiting time

PREVALENCE

Whereas incidence assesses the frequency of disease onset, prevalence is a measure of disease status: it deals with existing cases of disease. Prevalence is the proportion of people in a population having disease:

$$\text{Prevalence} = \frac{\text{Number of subjects having disease at a particular point in time}}{\text{Total number of subjects in the population}}$$

This definition of prevalence (synonym: prevalence proportion) reflects so-called 'point prevalence' that is most commonly used as the measure dealing with existing cases. Sometimes, however, researchers make use of 'period prevalence'. This measure includes existing cases at the start of the period plus the new cases that develop over the study period, for example, 1 year. This total is then divided by the total number of subjects in the population. Therefore, suppose that out of the 80 patients in our dialysis center there were 20 patients with ESRD due to glomerulonephritis at the beginning of 2005 and that during this year there were two new patients with the same condition taken into dialysis. Then the point prevalence of ESRD due to glomerulonephritis at the beginning of 2005 was $20/80 = 0.25$, but the period prevalence in the year 2005 was $22/80 = 0.275$.

Prevalence is affected by both the number of incident cases and the length of disease time. Given a steady state and a low prevalence, the prevalence equals the product of the incidence rate and the mean duration of disease (Prevalence = Incidence rate × Average disease duration). This is easily illustrated by the examples of aortic hemorrhage and ESRD. The first disease leads to rapid death resulting in a short disease duration; and therefore, its prevalence in the general population at any point in time will be extremely low. ESRD, on the other hand, has a relatively low incidence rate, but in comparison to aortic hemorrhage its survival is much higher, at least in developed countries. As a consequence, its average disease duration is much longer; and therefore, its prevalence is much higher compared to that of aortic hemorrhage. In the perspective of ESRD, this dependency of prevalence can further be illustrated by the fact that an increase in prevalence could be the result of a higher incidence of ESRD, of an improved survival or be the consequence of both.

Prevalence is said to reflect the burden of disease in a population. Burden can be measured in terms of costs, life expectancy, morbidity, quality of life, or other indicators. Knowledge of the burden of disease can help determine where investment in health should be targeted. Knowing that the average cost of dialysis patients per year within Medicare was 66 650 US dollars per patient in 2004,[6] multiplication by the number of prevalent patients will yield the total costs for dialysis patients within this part of the US health-care system. Example 3 is an illustration of how cost calculation per patient may be used to determine the economic burden of specific medication in dialysis patients.

Example 3 – Prevalence

Lorenzo et al.[7] did a study among 1312 hemodialysis patients from six centers in Spain, which was estimated to represent almost 10% of all hemodialysis patients in Spain. They performed a cost analysis to evaluate the economic burden of

mineral regulating therapy in this patient group. It turned out that on average the cost of this specific therapy was 1.68 Euro per patient per day.

Had the sample in the study of Lorenzo *et al.* indeed represented almost 10% of the hemodialysis patients in Spain and had it been a representative sample, the costs of mineral regulating therapy for this patient group in Spain could have been estimated at slightly more than 8 million Euro per year.

MEASURES OF EFFECT

Measures of disease occurrence may also be used to study the risk factors or causes of disease. To assess the effect of a risk factor, comparisons are made between, for example, disease frequency in people who have been exposed to that risk factor and disease frequency in those who have not been exposed to that risk factor. To express the size of such effects, one uses either differences between measures of disease occurrence or ratios of those measures. A later paper in this series will address these different measures of effect.

SUPPLEMENTARY MATERIAL

The supplementary material is available at www.kidney-international.org.

REFERENCES

1. http://www.cdc.gov/reproductivehealth/EpiGlossary/glossary.htm.
2. Puliyanda DP, Stablein DM, Dharnidharka VR. Younger age and antibody induction increase the risk for infection in pediatric renal transplantation: a NAPRTCS report. *Am J Transplant* 2007; **7**: 1–5.
3. Victora CG. What's the denominator? *Lancet* 1993; **342**: 97–99.
4. Rothman KJ. *Epidemiology: An Introduction*. Oxford University Press: New York, 2002.
5. Kramar R, Oberbauer R. Austrian Dialysis and Transplantation Registry (OEDTR), Annual Report 2005, Austrian Society of Nephrology (available at http://www.nephro.at/oedr2005/oedr2005.htm).
6. US Renal Data System. USRDS 2006 Annual Data Report: Atlas of End-Stage Renal Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases: Bethesda, MD, 2006.
7. Lorenzo V, Martin-Malo A, Perez-Garcia R *et al.* Prevalence, clinical correlates and therapy cost of mineral abnormalities among haemodialysis patients: a cross-sectional multicentre study. *Nephrol Dial Transplant* 2006; **21**: 459–465.